INTERNET-DRAFT
Intended Status: Experimental
J. Kumar
S. Anubolu
J. Lemon
Broadcom Inc.
H. Holbrook
Arista Networks
A. Ghanwani
Dell EMC
D. Cai
H. OU
AliBaba Inc.
Y. Li
Huawei
Expires: April 21, 2019 October 18, 2018

Inband Flow Analyzer
draft-kumar-ippm-ifa-00

Abstract

Inband Flow Analyzer (IFA) records flow specific information from an end station and/or switches across a network. This document discusses the method to collect data on a per hop basis across a network and perform localized or end to end analytics operations on the data. This document also describes a transport-agnostic header definition that may be used for tunneled and non-tunneled flows alike.

Status of this Memo

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Copyright and License Notice

Table of Contents

1. Introduction

This document describes an Inband Flow Analyzer (IFA) a mechanism to mark a packet to enable the collection of metadata regarding the analyzed flow. IFA defines an IFA header to mark the flow and direct the collection of analyzed metadata per marked packet per hop across a network. The ability to mark a packet using an IFA OAM header can be leveraged to create synthetic flows meant for network data collection. This document describes a mechanism that may be used to monitor live traffic and/or create synthetic flows. This document also describes IFA zones, IFA reports, and IFA metadata. IFA does not require changes to protocol headers in order to collect metadata or analyze flows. IFA puts minimal requirements on switching silicon.

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

IFA: Inband Flow Analyzer

MTU: Maximum Transmit Unit

1.2 Scope

This document describes IFA deployment, the type of traffic that is supported, header definitions, analytics, and data path functions.

IFA deployment involves defining an IFA zone and understanding the requirements in terms of traffic overhead and points of data collection. Given that IFA provides the ability to perform local analytics on the collected data, this document describes the scope of the analytics function as well. The scope of IFA is from an end station and/or ToR, through any/all nodes in the network, and terminating in a network switch and/or an end station.

IFA can create a synthetic stream of traffic and use it to collect metadata along the path. This sampled stream is later discarded. IFA can also insert metadata on a per packet basis in live traffic. Inband insertion of metadata can be done within the payload or via tail stamping. This draft defines an identification mechanism using a dedicated protocol type in the IP header for identifying IFA.

1.3 Applicability

IFA is capable of providing traffic analysis in an encapsulation-

agnostic manner. Simple TCP and UDP flows, as well as tunneled flows, can be monitored. IFA can be enabled on an end station, or it can be enabled just on network switches. Enabling IFA on an end station provides better scalability and visibility by monitoring intra end station or inter end station traffic. IFA performs best when there is hardware assistance for deriving the flow metadata in the data path. This document describes data path functions for IFA.

1.4 Motivation

The main motivation for IFA is to collect analyzed metadata on a per packet per flow basis for a given application. The IFA header is defined in order to work for any IP packet, and with minimal impact on hardware performance.

2. Requirements

IFA requirements are defined with operational efficiency, performance of the network, and cost of hardware in mind.

2.1 Encapsulation Requirements

IFA packets MUST be clearly marked and identifiable so that a networking element in the flow path can insert metadata or perform other IFA operations.

IFA packets need to be easily identified for performance reasons. IFA packet identification MUST be the same for all the encapsulation types. This means that expensive hardware modifications are not needed for supporting new protocol types.

Since IFA packet processing is a data path function, the IFA header MUST avoid next header chaining. Simple parsing in the switch hardware is required to maintain the switch performance and cost.

A single IFA encapsulation MUST support IPv4,IPv6 protocol types for tunneled and non-tunneled packets, preserving the fields used for load balancing hash computation.

IFA SHOULD support a checksum for the entire IFA metadata stack instead of per metadata element.

2.2 Operational Requirements

IFA MUST preserve the flow path across the network.

IFA MUST incur minimal traffic overhead.

IFA MUST provide an option to clone and truncate a packet to avoid disrupting the PMTU discovery of a network.

Cloning MUST be done at a sampled ratio to keep the network overhead minimum.

IFA MUST provide the ability to insert metadata on cloned traffic.

IFA MUST provide the ability to insert metadata on live traffic.

IFA MUST provide the ability to specify checksum validation on the IFA header and metadata.

IFA MUST provide the ability to define a zone using hop count.

IFA MUST provide the ability for a networking element to perform metadata insertion in the payload or append it via tail stamping.

IFA MUST be able to support an IFA zone name space, also referred to as a global name space.

IFA MUST be able to support a per hop name space, also referred to as a local name space.

2.3 Cost and Performance Requirements

The IFA header and metadata MUST be treated as foreign data present in the application data. IFA SHOULD be able to insert or strip the IFA data without modifying the layer 3 and layer 4 headers. This will help keep the cost of hardware down with no degradation in performance.

IFA MUST support the ability to clone and/or truncate live traffic for IFA metadata insertion. This is needed for PMTU protocols to work well within the IFA zone.

The IFA header MUST provide the ability to differentiate between a cloned packet and an original packet. This is needed for hardware to be able to identify and filter the cloned traffic at the edge of an IFA zone.

IFA encapsulation MUST NOT impact the parse depth of hardware for packet processing. This will help avoid degradation in performance when IFA is enabled.

IFA MUST NOT require pre-allocation for reserving the space in a packet. The overhead of managing reserved space in a packet can result in performance degradation.

3. IFA Operations

IFA performs flow analysis, and possible actions on the flow data, inband. Once a flow is enabled for analysis, a node with the role of "Initiator" makes a copy of the flow or samples the live traffic flow, or tags a live traffic flow for analysis and data collection. Copying of a flow is done by sampling or cloning the flow. These new packets are representative packets of the original flow and possess the exact same characteristics as the original flow. This means that IFA packets traverse the same path in the network and same queues in the networking element as the original packet would. Figure 1 shows the IFA based Telemetry Framework. The terminating node is responsible for terminating the IFA flow by summarizing the metadata of the entire path and sending it to a Collector.

```
                            +----------+
                            |          |
                            |Collector |--------------+
                            |          |              |
                            +----------+              |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
     +---------------+       +-------------+       +----+-----------+
     |Initiator Node|       |Transit Node|       |Terminating Node|
     |   +-------+   |       |  +-------+  |       |    +-------+   |
     |   | IFA   |   |------->|  | IFA   |  |------->|    | IFA   |   |
     |   +-------+   |IFA flow|  +-------+  |IFA flow|    +-------+   |
     +---------------+       +-------------+       +----------------+
```

Figure 1 IFA Zone Framework

3.1 IFA Zones

An IFA zone is the domain of interest where IFA monitoring is enabled. An IFA zone MUST have designated IFA function nodes. An IFA zone can also be controlled by setting an appropriate TTL value in the L3 header. Initiating and Terminating function nodes are always at the edge of the IFA zone. Internal nodes in the IFA zone are always Transit function nodes.

3.2 IFA Function Nodes

There are three types of IFA functional nodes.

3.2.1 Initiating Function Node

An end station, a switch, or any other middlebox can perform IFA initiating function. It is advantageous to keep this role closest to the application to maximize flow visibility. An IFA initiating function node performs the following functions:

- Samples the flow traffic of interest based on a configuration.
- Converts the traffic into an IFA flow by adding an IFA header to each sample.
- Updates the packet with initiating function node metadata.
- May mandate a specific template ID metadata by all networking elements.
- May mandate tail stamping of metadata by all networking elements.

3.2.2. Transit Function Node

An IFA transit node is responsible for inserting transit node metadata in an IFA packet.

3.2.3. Terminating Function Node

An IFA terminating node is responsible for the following:
- Inserts terminating node metadata in an IFA packet.
- Performs a local analytics function on one or more segments of metadata, e.g., threshold breach for residence time, congestion notifications, and so on.
- Filters an IFA flow in case of cloned traffic
- Removes the IFA headers and forwards the packet in case of live traffic

3.3 IFA Cloning, Truncation, and Drop

IFA allows cloning live traffic. Cloned traffic will have same network path characterization as the original traffic.

Cloned traffic can be truncated to accommodate for PMTU of the IFA zone.

Clone traffic MUST be dropped by the terminating node of IFA zone.

3.4 IFA Header

A compact IFA header is described below. An experimental IP protocol number is used in the IP header to identify an IFA packet. The IP header protocol type field is copied into the IFA header NextHdr

field for hardware to correctly interpret the layer 4 header.

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       IPv4 Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Version|  IHL  |Type of Service|          Total Length         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |         Identification        |Flags|      Fragment Offset    |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |  Time to Live | Protocol = IFA|         Header Checksum       |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                       Source IPv4 Address                     |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                    Destination IPv4 Address                   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    or
       IPv6 Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Version| Traffic Class |           Flow Label                  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |         Payload Length        |Next Hdr = IFA |   Hop Limit   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                                                               |
       +                                                               +
       |                                                               |
       +                         Source IPv6 Address                   +
       |                                                               |
       +                                                               +
       |                                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                                                               |
       +                                                               +
       |                                                               |
       +                      Destination IPv6 Address                 +
       |                                                               |
       +                                                               +
       |                                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

       IFA Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Ver=2.0|  GNS  |NextHdr = IP_xx|      Flags    |  Curr Length  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2 IFA 2.0 Header Format

(1) Version (4 bits) - Specifies the version of IFA header.
(2) GNS (4 bits) - Global Name Space. Specifies the IFA zone scoped name space for IFA metadata.
(3) Protocol Type (8 bits) - IP Header protocol type. This is copied from the IP header.
(4) Flags (8 bits)

0: Option1 - IFA Option 1. Indicates the presence of the Option 1 header.
1: TA - Turn Around. Indicates that the IFA packet needs to be turned around at the terminating node of the IFA zone.
2: I - Inband. Indicates this is live traffic. Strip and forward MUST be performed by the terminator node if this bit is set.
3: TS - Tail Stamp. Indicates the IFA zone is requiring tail stamping of metadata.
4: Reserved. MUST be initialized to 0 on transmission and ignored on receipt.
5: Reserved. MUST be initialized to 0 on transmission and ignored on receipt.
6: Reserved. MUST be initialized to 0 on transmission and ignored on receipt.
7: Option2 - IFA Option 2. Indicates the presence of the Option 2 checksum header. The checksum MUST be computed and updated for the IFA header and metadata.

(5) Current Length (8 bits) - Specifies the current length of the metadata in multiples of 4 octets.

3.4.1 IFA Option 1 Header

This header is optional.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Request Vector| Action Vector |   Hop Limit   |   Max Length  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3 IFA Option 1 Header Format

Request Vector (8 bits) - This vector specifies the presence of fields as specified by GNS. Fields are always quad octet aligned. This field can be made extensible by defining a new GNS for an IFA zone.

Action Vector (8 bits) - This vector specifies node-local or end-to-

end action on the IFA packets.

Hop Limit (8 bits) - Specifies the maximum allowed hops in an IFA zone. This field is initialized by the initiator node. The hop limit MUST be decremented at each hop. If the hop limit becomes 0, transit nodes MUST stop inserting metadata. A value of 0xFF means that the Hop limit check MUST be ignored.

Max Length (8 bits) - Specifies the maximum allowed length of metadata stack in multiples of 4 octets. This field is initialized by the initiator node. Each node in the path MUST compare the current length with the max length, and if the current length equals or exceeds the max length, the transit nodes MUST stop inserting metadata.

3.4.2 IFA Option 2 Header

This header is optional.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Checksum             |           Reserved            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4 IFA Option 2 Header Format

Checksum (16 bits) - The checksum covering the IFA header and metadata stack.

Reserved (16 bits) - Reserved. MUST be initialized to 0 on transmission and ignored on receipt.

3.5 IFA Metadata

This is the information inserted by each hop after the IFA header. IFA metadata can be inserted at the following offsets:

- Payload Stamping: Immediately after the layer 4 header
- Tail Stamping: After the end of the packet

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  LNS  |                     Device ID                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                                                               |
~                LNS defined metadata (contd)                   ~
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3 IFA Metadata Format

The IFA metadata header contains a set fields as defined by the name space identifier. Two types of name space identifiers are proposed.

3.5.1 Global Name Space (GNS) Identifier

A GNS is specified in the IFA header by the initiator node. The scope of a GNS is an IFA zone. All networking elements in an IFA zone MUST insert metadata as per the GNS ID specified in the IFA header.

This is defined as the "Uniform Mode" of deployment.

A GNS value of 0xF indicates that metadata in an IFA zone is defined by the LNS of each hop.

The advantage of using the uniform mode is having a simple and uniform metadata stack. This means less load on a collector for parsing.

The disadvantage is that metadata fields are supported based on the least capable networking element in the IFA zone.

3.5.2 Local Name Space (LNS) Identifier

An LNS is specified in the metadata header. A GNS value of 0xF in the IFA header indicates the presence of an LNS.

A switch pipeline MUST parse the GNS field in the IFA header. The parsing result will dictate the name space ID that the hop needs to comply with.

This is defined as the "Non-uniform Mode" of deployment.

The advantage of using the non-uniform mode is having a flexible

metadata stack. This allows each hop to include the most relevant data for that hop.

The disadvantage is more complex parsing by a collector.

3.5.3 Device ID

A 28-bit unique identifier for the device inserting the metadata. If a GNS other than 0xF is present, then the device ID can be expanded to a 32 bit value. This is to support including an IPv4 loopback address as a Device ID.

3.6 IFA Network Overhead

A common problem associated with inserting metadata on a per packet per flow basis is the amount of traffic overhead on the network. IFA 2.0 is defined to minimize the overhead on the network.

```
IFA Base Header: 4 octets
IFA Option 1   : 4 octets
IFA Option 2   : 4 octets

IFA metadata with LNS: 4 octets
IFA metadata with GNS: 0 octets

Minimum Overhead:
   IFA header    : 4 octets
   IFA Metadata  : 4 octets
   Total Overhead: 8 octets per packet
```

3.7 IFA Analytics

There are two kinds of actions considered in this proposal.

(1) Action Bit MAP in IFA Header - This is encoded in the IFA header. Each node in the path MAY use the action bitmap to insert or not insert the metadata based on exceeding a locally-specified threshold. Not inserting the metadata is indicated by setting the field value to -1 (all 1s).

(2) Terminating Node Actions - A terminating node may decide to perform threshold or other actions on the set of metadata in the packet. This information is not encoded in the IFA header.

3.8 IFA Packet Format

The IFA header is treated as a layer 3 header extension.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                         IP Header                             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                        IFA Header                             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                      Layer 4 Header                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                     IFA Metadata Stack                        |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ~                                                               ~
 |                      Layer 4 Header                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                 Figure 4 IFA Packet Format
```

3.8.1 TCP/UDP Packet

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       IPv4 Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Version|  IHL  |Type of Service|          Total Length         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |         Identification        |Flags|      Fragment Offset    |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |  Time to Live | Protocol = IFA|         Header Checksum       |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                       Source IPv4 Address                     |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                    Destination IPv4 Address                   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    or
       IPv6 Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Version| Traffic Class |           Flow Label                  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |         Payload Length        |Next Hdr = IFA |   Hop Limit   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                                                               |
       +                                                               +
       |                                                               |
       +                      Source IPv6 Address                      +
       |                                                               |
       +                                                               +
       |                                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                                                               |
       +                                                               +
       |                                                               |
       +                   Destination IPv6 Address                    +
       |                                                               |
       +                                                               +
       |                                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

       IFA Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Ver=2.0|  GNS  |NextHdr=TCP/UDP|      Flags    |  Curr Length  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

       TCP Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |          Source Port          |       Destination Port        |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                        Sequence Number                        |
```

```
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                    Acknowledgment Number                      |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Data |           |U|A|P|R|S|F|                               |
      | Offset| Reserved  |R|C|S|S|Y|I|            Window             |
      |       |           |G|K|H|T|N|N|                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |           Checksum            |         Urgent Pointer        |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                    Options                    |    Padding    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      IFA Metadata Stack:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                    IFA Metadata Stack                         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      TCP Payload:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                             data                              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5 TCP/UDP IFA Packet Format

3.8.2 VxLAN Packet

```
       0                   1                   2                   3
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      IPv4 Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Version|  IHL  |Type of Service|          Total Length         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |         Identification        |Flags|      Fragment Offset    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Time to Live | Protocol = IFA|         Header Checksum       |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                       Source IPv4 Address                     |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                    Destination IPv4 Address                   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   or
      IPv6 Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Version| Traffic Class |           Flow Label                  |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |         Payload Length        |Next Hdr = IFA |   Hop Limit   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                                                               |
      +                                                               +
      |                                                               |
      +                     Source IPv4 Address                       +
      |                                                               |
      +                                                               +
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                                                               |
      +                                                               +
      |                                                               |
      +                  Destination IPv6 Address                     +
      |                                                               |
      +                                                               +
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      IFA Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ver=2.0|  GNS  | NextHdr = UDP |      Flags    |  Curr Length  |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      Outer UDP Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |           Source Port         |       Dest Port = VXLAN Port  |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |           UDP Length          |        UDP Checksum           |
```

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

   IFA Metadata Stack:
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     IFA Metadata Stack                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

   VXLAN Header:
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |R|R|R|R|I|R|R|R|            Reserved                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                VXLAN Network Identifier (VNI) |   Reserved    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                             data                              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6 VxLAN IFA Packet Format

3.8.3 GRE Packet

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       IPv4 Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Version|  IHL  |Type of Service|          Total Length         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |         Identification        |Flags|      Fragment Offset    |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |  Time to Live | Protocol = IFA|         Header Checksum       |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                       Source IPv4 Address                     |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                    Destination IPv4 Address                   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    or
       IPv6 Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Version| Traffic Class |           Flow Label                  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |         Payload Length        |Next Hdr = IFA |   Hop Limit   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                                                               |
       +                                                               +
       |                                                               |
       +                     Source IPv6 Address                       +
       |                                                               |
       +                                                               +
       |                                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                                                               |
       +                                                               +
       |                                                               |
       +                  Destination IPv6 Address                     +
       |                                                               |
       +                                                               +
       |                                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

       IFA Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |Ver=2.0|  GNS  | NextHdr = GRE |      Flags    |  Curr Length  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

       GRE Header:
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |C|       Reserved0       | Ver |         Protocol Type         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |      Checksum (optional)      |       Reserved1 (Optional)    |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

IFA Metadata Stack:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     IFA Metadata Stack                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

GRE Payload:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              data                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
               Figure 7 GRE IFA Packet Format
```

3.8.4 Geneve Packet

```
       0                   1                   2                   3
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      IPv4 Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Version|  IHL  |Type of Service|          Total Length         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |         Identification        |Flags|      Fragment Offset    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Time to Live | Protocol = IFA|         Header Checksum       |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                       Source IPv4 Address                     |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                    Destination IPv4 Address                   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   or
      IPv6 Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Version| Traffic Class |           Flow Label                  |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |         Payload Length        |Next Hdr = IFA |   Hop Limit   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                                                               |
      +                                                               +
      |                                                               |
      +                     Source IPv6 Address                       +
      |                                                               |
      +                                                               +
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                                                               |
      +                                                               +
      |                                                               |
      +                  Destination IPv6 Address                     +
      |                                                               |
      +                                                               +
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      IFA Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ver=2.0|  GNS  | NextHdr = UDP |      Flags    |  Curr Length  |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      Outer UDP Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |       Source Port = xxxx      |    Dest Port = Geneve Port    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |           UDP Length          |        UDP Checksum           |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

IFA Metadata Stack:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     IFA Metadata Stack                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Geneve Header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Ver|  Opt Len  |O|C|    Rsvd.  |          Protocol Type        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Virtual Network Identifier (VNI)       |    Reserved   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Variable Length Options                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             data                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8 Geneve IFA Packet Format

3.9 IFA Load Balancing

IFA changes the IP protocol field value to IFA protocol number. IP protocol field value is included in the hash computation. This will impact load balancing of flows.

Forwarding plane MUST support reading the IP protocol field value stored in IFA NextHDR field for hash computation.

4. Interoperability Considerations

Version 2.0 of this protocol specification is not backward compatible with version 1.0.

5. Security Considerations

A successful attack on an OAM protocol can prevent the detection of failures or anomalies, or create a false illusion of nonexistent ones.

The metadata elements of IFA can be used by attackers to collect information about the network hops.

Adding IFA headers or adding to IFA metadata can be used to consume resources within the path being monitored or by a collector.

Adding IFA headers or adding to IFA metadata can be used to force

exceeding the MTU for the path being monitored resulting in fragmentation and/or packet drops.

IFA is expected to be deployed within controlled network domains, containing attacks to that controlled domain. Limiting or preventing monitoring or attacks using IFA requires limiting or preventing unauthorized access to the domain in which IFA is to be used, and preventing leaking IFA metadata beyond the controlled domain.

6. References

6.1 Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

6.2 Informative References

[RFC791] https://tools.ietf.org/html/rfc791

[RFC6864] https://tools.ietf.org/html/rfc6864

[RFC3514] https://tools.ietf.org/html/rfc3514

[IFA 1.0] https://tools.ietf.org/html/draft-kumar-ifa-00

Authors' Addresses

Jai Kumar
Broadcom Inc.
Email: jai.kumar@broadcom.com

Surendra Anubolu
Broadcom Inc.
Email: surendra.anubolu@broadcom.com

John Lemon
Broadcom Inc.
Email: john.lemon@broadcom.com

Hugh Holbrook
Arista Networks
Email: holbrook@arista.com

Anoop Ghanwani

Dell EMC
Email: anoop.ghanwani@dell.com

Dezhong Cai
AliBaba Inc.
Email: d.cai@alibaba-inc.com

Heidi OU
AliBaba Inc.
Email: heidi.ou@alibaba-inc.com

Yizhou Li
Huawei Technologies
EMail: liyizhou@huawei.com

Appendix A

Appendix A is just for informational purposes. The following options were considered for the IFA protocol.

A.1 Probe Marker

One of the challenges of using probe signatures in an IFA header is a false positive.

The IFA version 2.0 header takes care of large header sizes and identification based on probe markers. Probe markers can cause false positives if there is a match on the first 64 bits of the layer 4 payload.

This approach is not a preferred approach, but is supported by this draft as a version 1.0 header.

A.2 DSCP

[RFC791] EXP/LU Pool 3 can be used for identifying IFA packets. CU bits can be used for identifying IFA packets.

The problem with using TOS bits is that they are pervasively used in the network deployment and are responsible for affecting the forwarding decision.

This approach is not supported or recommended by this draft.

A.3 IP Options

[RFC791] The Options provide for control functions needed or useful in some situations but unnecessary for the most common communications. The options include provisions for timestamps, security, and special routing.

There are various problems with this approach.
(1) The IPv4 header size can become arbitrarily large with the presence of options.
(2) A switch pipeline typically handles IP option packets as exception path processing and punts them to a host CPU.
(3) IP options make the construction of firewalls cumbersome, and are typically disallowed or stripped at the perimeter of enterprise networks by firewalls.

This approach is not supported or recommended by this draft.

A.4 IPv4 Identification or Reserved Flag

[RFC6864] [RFC3514] Another suggestion is to use the IPv4 identification field or reserved flag. This suggestion is also discarded and not supported for the following reasons:

[RFC6864] prohibits usage of id field for any other purposes.

[RFC3514] prohibits using flags bit 0 for security reasons.